Development of Logistic Regression Models for Portland Harbor

Logistic regression models (LRMs) were developed from Portland Harbor data for 293 samples with matching sediment chemistry and toxicity (US EPA 2005; Field et al. 2002; Field et al 1999). Individual LRMs were developed for each species. The individual models were combined into a single model that uses the maximum probability (pMax) of the selected individual models for each sample. The methods used to develop the individual and combined models are described below.

Data Treatment

Chemistry

Chemicals with at least 50 detected values were included in the individual model development (63 chemicals, including summed parameters; Table 1). Individual PCB congeners, dioxin/furans, and most sediment conventionals (except for ammonia and sulfides) were excluded. Perylene, dibenzothiophene, and 1-methylnaphthalene were excluded from model selection because these chemicals were only analyzed in the Phase 3 data collection (n=60).

LRMs were developed for individual PAHs and DDTs and their sums. Because the models were developed and evaluated independently, the redundancy was not considered to be a problem. For samples with different PAH compositions or incomplete analyte list, having both individual and summed parameters may be beneficial.

Below detection values were excluded from the data used for individual model development.

In addition to dry weight concentrations, models were developed for concentrations normalized to Total Organic Carbon (TOC), proportion fines (concentration*proportion fines), and a combination of TOC and fines [(concentration* proportion fines) /(fraction OC)]. TOC ranged between 0.1% and 15.2%. No adjustments to TOC or percent fines were used in the calculated values. The normalizations were applied in the same manner to all chemicals.

Toxicity Endpoints

Toxicity for each species (*Hyalella* and *Chironomus*) and endpoint (survival and biomass) was classified into 4 levels according to EPA's reference envelope approach. Individual models were developed for each species. The highest toxicity level for either the survival and biomass endpoint for each species was used in species-specific model development. The highest toxicity level from either species was used in the combined model development.

Individual Logistic Regression Models

Individual logistic regression models were developed for each chemical (Field et al 1999; Field et al 2002; EPA 2005). Models were derived using a binary toxic/non-toxic designation according to the toxicity level.

Toxicity Classification
Individual logistic regression models were developed for each species for all 3 toxicity levels.

Screening
Toxic samples were excluded from individual model development using 3 screening approaches:

- 1X: if concentration of a toxic sample was less than or equal to the arithmetic mean of the non-toxic samples, the sample was excluded from individual model development for that chemical;
- 2X: if concentration of a toxic sample was less than or equal to twice the arithmetic mean of the non-toxic samples, the sample was excluded from individual model development for that chemical;
- 2G: if concentration of a toxic sample was less than or equal to twice the geometric mean of the non-toxic samples, the sample was excluded from individual model development for that chemical

Normalization
Individual models were developed for each chemical for the 4 normalization approaches described above.

Model development
Using only Portland Harbor data, 36 individual logistic regression models were developed for each species (3 toxicity levels, 3 screening levels, and 4 normalizations). The models estimate the probability that a sample will be toxic at the defined toxicity level from the log (base 10) chemical concentration. Models with a normalized chi square of greater than 0.15 were considered acceptable for inclusion in the combined models. Individual models also were required to have a ratio of T75/T50 greater than 1.25 to ensure a gradient in the model.

Combined Model

Individual models for each species were combined into an all endpoint model using an individual chemical model selection approach. The combined model (pMAX model) uses the maximum probability for each sample from the probabilities from the selected individual chemical models. All individual models were evaluated and selected based on toxicity greater than or equal to toxicity level 2 in any one of the 4 toxicity endpoints.

Treatment of below detection limit (BDL) values in combined model: BDL values with ½ BDL probability < 0.25 included in combined model evaluation at ½ BDL. If ½ BDL value had a probability > 0.25, the result was excluded from the combined model calculation of the maximum probability (pMAX) for the sample.

Each individual model was applied to the entire dataset (293 samples) and several metrics were calculated for use in model selection:

- Positive likelihood ratio based on a probability threshold of 0.5.
- Proportion toxic for samples with a probability >= 0.5 (hit reliability for a probability threshold of p=0.5).

- Number of samples correctly predicted as toxic for samples with a probability > 0.5.
- Proportion toxic for samples with a probabilities between 0.5 and 0.75 and the proportion toxic for samples >= 0.75.

The selected models and parameters are shown in Table 2 and the proportion of samples toxic for the models by probability quartile is shown in Table 3.

The individual model with highest positive likelihood ratio for a chemical was the selected model for 31 of the 40 models (Table 4). Of the remaining 9 models, 3 had the highest or second highest hit reliability for a probability threshold of p>=0.5 and 6 were selected based on the number of samples correctly classified also having a high hit reliability (proportion toxic>=0.75).

The combined models were optimized for hit reliability and the number of samples correctly classified as hits by eliminating models that adversely affected overall combined model performance. Calibration of the combined pMAX models was based on the performance of the individual chemical models where the individual model had the highest probability for a sample. Individual models with a high proportion of false positives were removed from the combined pMax model. Chemicals with the highest probability for a sample (pMAX value) and a low proportion of samples toxic for probability > 0.5 were removed (typically 2-3 models at a time in an iterative process) until all remaining models meet performance standards.

Table 5 shows the final selected individual models. All of the selected model chemicals with a pMAX value of greater than 0.5, correctly predicted at least one sample. For all of the models except PCBs, the proportion toxic was greater than or equal to 0.5. Half (20) of the selected models did not have the maximum probability for a sample for a pMAX of greater than 0.5. These models could be removed from the combined model without affecting overall model performance for the 293 sample derivation database, but may be useful for samples with incomplete chemistry.

Reliability:
The reliability statistics (Shepard 2010) for the final combined model are shown in Table 6. For toxicity level 2, pMax thresholds between 0.50 and 0.52 have false negative values of <= 0.5, false positive values of <= 0.1, and overall correct classification rates of >= 0.75. For toxicity level 3, pMax thresholds between 0.59 and 0.72 have false negative values of <= 0.5, false positive values of <= 0.1, and overall correct classification rates of >= 0.75.

Model performance relative to percent fines
Because of the normalization approach that included an adjustment for percent fines in some of the selected individual chemical models, the question was raised whether the model performance was only a reflection of the correlation between fine-grained sediments and toxicity. Of the 293 samples, 54 had percent fines greater 75% and 51.9% (28) of those samples had toxicity level of 2 or greater in at least one endpoint. Of the high fines samples, 18 had a pMax of greater than 0.5 and 72.2% of those samples were toxic (level 2 or greater) and 81.8% of 11 samples with pMax > 0.75. This suggests that the pMax model, which includes combinations of chemistry

and fines, provides additional information that can be used to predict toxicity from sediment chemistry.

Outline of LRM Approach

Individual chemical-specific model development

- Establish list of chemicals for potential model development
- Establish toxicity levels according to EPA reference envelope approach
- Use the highest response for each species (survival or biomass) as the toxicity level for model development.
- Derive individual models for each chemical (by species, 4 normalizations, 3 toxicity levels, and 3 screening approaches) excluding below detection limit samples.
- Include models in the candidate list for combined model selection if they met all of the following criteria:
  - normalized chi-square value > 0.15,
  - ratio of T75/T50>1.25,
  - at least than 50 samples in the screened model data set,
  - the percent of samples toxic greater than 50% for probability > 0.5

Individual model selection

- Models for each chemical were ranked by positive likelihood ratio (using probability threshold of 0.5), the proportion of samples toxic for probability > 0.5, and the number of samples correctly classified as toxic for probability > 0.5.
- Models with the highest positive likelihood ratio for each chemical were selected as first step
- Models with high proportion of samples toxic for probability > 0.5 and a greater number of samples correctly classified may be selected instead of the model with highest likelihood ratio (BPJ)

Combined model calibration: removal of individual chemical models with low hit reliability in the pMAX model

- Chemicals that set the pMAX value for a sample with low proportion of samples toxic for probability > 0.5 were removed (typically 2-3 models at a time in an iterative process) until all remaining models meet performance standards.

References

Field LJ, MacDonald DD, Norton SB, Ingersoll CG, Severn CG, Smorong D, Lindskoog R. 2002. Predicting amphipod toxicity from sediment chemistry using logistic regression models. Environ. Toxicol. Chem. 21: 1993-2005.

Field LJ, MacDonald DD, Norton SB, Severn CG, Ingersoll CG. 1999. Evaluating sediment chemistry and toxicity data using logistic regression modeling. Environ. Toxicol. Chem. 18:1311-1322.

Shepard B. 2010. Reliability statistics calculation spreadsheet.

U.S. EPA (2005). Predicting toxicity to amphipods from sediment chemistry. National Center for Environmental Assessment, Washington, DC; EPA/600/R-04/030.